

**PENGELOMPOKAN DAERAH TERDAMPAK COVID-19 DI PROVINSI
JAWA TIMUR MENGGUNAKAN METODE *HIERARCHICAL K-MEANS*
*CLUSTERING (HK-MEANS)***

SKRIPSI



Disusun Oleh
MU'ASATUDZ DZIKRIYYAH
H72217053

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2021

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : MU'ASATUDZ DZIKRIYYAH

NIM : H72217053

Program Studi : Matematika

Angkatan : 2017

Mengatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "PENGELOMPOKAN DAERAH TERDAMPAK COVID-19 DI PROVINSI JAWA TIMUR MENGGUNAKAN METODE *HIERARCHICAL K-MEANS CLUSTERING (HK-MEANS)*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian inisaya buat dengan sebenar-benarnya.

Surabaya, 17 Juni 2021

Yang menyatakan,



MU'ASATUDZ DZIKRIYYAH

NIM.H72217053

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : MU'ASATUDZ DZIKRIYYAH
NIM : H72217053
Judul Skripsi : PENGELOMPOKAN DAERAH TERDAMPAK COVID-19 DI PROVINSI JAWA TIMUR MENGGUNAKAN METODE *HIERARCHICAL K-MEANS CLUSTERING (HK-MEANS)*

telah diperiksa dan disetujui untuk diujikan.

Surabaya, 17 Juni 2021

Pembimbing I



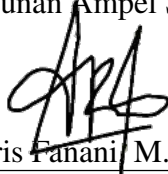
Nurissaidah Minnuha, M.Kom
NIP. 199011022014032004

Pembimbing II



Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Aris Fanani, M.Kom
NIP. 198701272014031002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : MU'ASATUDZ DZIKRIYYAH
NIM : H72217053
Judul Skripsi : PENGELOMPOKAN DAERAH TERDAMPAK COVID-19 DI PROVINSI JAWA TIMUR MENGGUNAKAN METODE *HIERARCHICAL K-MEANS CLUSTERING (HK-MEANS)*

Telah dipertahankan di depan Tim Penguji
pada tanggal 17 Juni 2021

Mengesahkan,
Tim Penguji

Penguji I

Nurissaidah Ulimnuha, M.Kom
NIP. 199011022014032004

Penguji II

Lutfi Hakim, M.Ag
NIP. 197312252006041001

Penguji III

Dian Candra Rini Novitasari, M.Kom
NIP. 198511242014032001

Penguji IV

Yuniar Farida, ST
NIP. 19790527014032002

Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Dr. Hj. Ev. Ratmatur Rusydiyah, M.Ag
NIP. 197312272005012003



**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax. 031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : MU'ASATUD2 DZIKRIYYAH
NIM : H72217053
Fakultas/Jurusan : SAINTEK / MATEMATIKA
E-mail address : muasatud2dzg8@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

PENGELOMPOKAN DAERAH TERDAMPAK COVID-19 DI PROVINSI
JAWA TIMUR MENGGUNAKAN METODE HIERARCHICAL K-MEANS
CLUSTERING (HK-MEANS)

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 16 Juli 2021

Penulis

(MU'ASATUD2 D.)

CLUSTERING (HK-MEANS)

19 sekarang menjadi sebuah pandemi yang terjadi di seluruh dunia. Khususnya di provinsi Jawa Timur yang merupakan salah satu provinsi di Indonesia yang penyebaran Covid-19 juga terus meningkat. Untuk itu, penurunan sampai saat ini. Tingginya penyebaran Covid-19 ini, maka perlu adanya pengelompokan daerah terdampak Covid-19 untuk mengetahui kemiripan karakteristik atau kriteria yang dapat dilakukan dalam pengelompokan ini menggunakan analisis *cluster*. Penelitian ini menggunakan Metode *Clustering (HK-Means)* dengan Studi Kasus : Jawa Timur. Metode *Means Clustering (HK-Means)* merupakan gabungan dari *Hierarchical Clustering* yaitu metode (*Single Linkage, Complete Linkage*) dengan *K-means*, digabungkan agar dapat meningkatkan kualitas *cluster* dengan kualitas yang lebih baik dan tinggi. Kelebihan yaitu pada penentuan pusat awal *cluster*. Untuk membandingkan dengan metode *Hierarchical Clustering* untuk mendapatkan *cluster* yang lebih baik. Ukuran jarak yang digunakan adalah *euclidean*. Hasil analisis *cluster* yang terbaik adalah

Covid-19 sekarang menjadi sebuah pandemi yang terjadi di seluruh dunia. Khususnya di provinsi Jawa Timur yang merupakan salah satu provinsi di Indonesia yang penyebaran Covid-19 juga terus meningkat. Penurunan sampai saat ini. Tingginya penyebaran Covid-19 ini maka perlu adanya pengelompokan daerah terdampak Covid-19 untuk mengetahui kemiripan karakteristik atau kriteria yang dapat dilakukan dalam pengelompokan ini menggunakan analisis *cluster*. Penelitian ini menggunakan Metode *Mean* (*HK-Means*) dengan Studi Kasus : Jawa Timur. Metode *Mean Clustering (HK-Means)* merupakan gabungan dari *Mean Clustering* yaitu metode (*Single Linkage, Complete Linkage*) dengan *K-means*, digabungkan agar dapat meningkatkan kualitas *cluster* dengan kualitas yang lebih baik dan tinggi. Kelebihan yaitu pada penentuan pusat awal *cluster*. Untuk membandingkan dengan metode *Hierarchical Clustering* untuk mendapatkan *cluster* yang lebih baik. Ukuran jarak yang digunakan adalah *euclidean*. Hasil analisis *cluster* yang terbaik adalah

Covid-19 sekarang menjadi sebuah pandemi yang terjadi di seluruh dunia. Khususnya di provinsi Jawa Timur yang merupakan salah satu provinsi di Indonesia yang penyebaran Covid-19 juga terus meningkat. Penurunan sampai saat ini. Tingginya penyebaran Covid-19 ini maka perlu adanya pengelompokan daerah berdasarkan tingkat penyebaran untuk mengetahui kemiripan karakteristik atau kriteria yang dapat dilakukan dalam pengelompokan ini menggunakan analisis *cluster*. Penelitian ini menggunakan Metode *Mean* (*HK-Means*) dengan Studi Kasus : Jawa Timur. Metode *Mean Clustering (HK-Means)* merupakan gabungan dari *Mean Clustering* yaitu metode (*Single Linkage, Complete Linkage*) dengan *K-means*, digabungkan agar dapat meningkatkan kualitas *cluster* dengan kualitas yang lebih baik dan tinggi. Kelebihan yaitu pada penentuan pusat awal *cluster*. Untuk membandingkan dengan metode *Hierarchical Clustering* untuk mendapatkan *cluster* yang lebih baik. Ukuran jarak yang digunakan adalah *euclidean*. Hasil analisis *cluster* yang terbaik adalah

Covid-19 sekarang menjadi sebuah pandemi yang terjadi di seluruh dunia. Khususnya di provinsi Jawa Timur yang merupakan salah satu provinsi di Indonesia yang penyebaran Covid-19 juga terus meningkat. Penurunan sampai saat ini. Tingginya penyebaran Covid-19 ini maka perlu adanya pengelompokan daerah berdasarkan tingkat penyebaran untuk mengetahui kemiripan karakteristik atau kriteria yang dapat dilakukan dalam pengelompokan ini menggunakan analisis *cluster*. Penelitian ini menggunakan Metode *Mean* (*HK-Means*) dengan Studi Kasus : Jawa Timur. Metode *Mean Clustering (HK-Means)* merupakan gabungan dari *Mean Clustering* yaitu metode (*Single Linkage, Complete Linkage*) dengan *K-means*, digabungkan agar dapat meningkatkan kualitas *cluster* dengan kualitas yang lebih baik dan tinggi. Kelebihan yaitu pada penentuan pusat awal *cluster*. Untuk membandingkan dengan metode *Hierarchical Clustering* untuk mendapatkan *cluster* yang lebih baik. Ukuran jarak yang digunakan adalah *euclidean*. Hasil analisis *cluster* yang terbaik adalah

GROUPING AREAS AFFECTED BY COVID-19 IN EAST JAVA PROVINCE USING *HIERARCHICAL K-MEANS CLUSTERING* (*HK-MEANS*) METHODS

Keywords: Covid-19, East Java, *Cluster Analysis*, *HK-Means*, and *Silhouette Coefficients*

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN	ii
LEMBAR PERSETUJUAN PEMBIMBING	iii
PENGESAHAN TIM PENGUJI SKRIPSI	iv
LEMBAR PERSETUJUAN PUBLIKASI	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG	xiv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	7
1.3. Tujuan Penelitian	7
1.4. Manfaat Penelitian	8
1.5. Batasan Masalah	8
1.6. Sistematika Penulisan	9
II TINJAUAN PUSTAKA	10
2.1. Covid-19 di Jawa Timur	10
2.2. Standarisasi Data	12
2.3. Uji Multikolineritas	12
2.4. Jarak <i>Euclidean</i>	14
2.5. <i>Clustering</i>	14
2.6. <i>Hierarchical Clustering</i>	17

2.6.1. Metode <i>agglomerative</i> (penggabungan)	18
2.6.2. Metode <i>Divisive</i> (pemecahan)	20
2.7. <i>K-Means</i>	20
2.8. Metode <i>Hierarchical K-Means Clustering</i>	22
2.9. Pengujian Metode Terbaik dengan <i>Silhouette coefficient</i>	24
2.10. Kewajiban Menjaga Kesehatan dan Keselamatan Jiwa	26
2.10.1. Memakai Masker	27
2.10.2. Selalu Mencuci tangan	28
2.10.3. Tetap berada di rumah	29
2.10.4. Tidak Berprasangka Buruk Kepada Allah Ta'ala	30
2.10.5. Bersikap Optimis dan Berucap yang Baik	31
III METODE PENELITIAN	32
3.1. Jenis Penelitian	32
3.2. Sumber Data	32
3.3. Variabel Penelitian	32
3.4. Tahapan Penelitian	33
IV HASIL DAN PEMBAHASAN	37
4.1. Deskripsi Data	37
4.2. Standarisasi Data	39
4.3. Uji Multikolineritas	39
4.4. Proses <i>Hierarchical K-Means Clustering</i>	44
4.4.1. <i>Hierarchical Clustering</i>	44
4.4.2. Menentukan Pusat <i>Cluster (Centroid)</i>	54
4.4.3. <i>K-Means</i>	57
4.5. Pemilihan Metode Terbaik Dengan <i>Sillhoutte Coefficient</i>	61
4.6. Interpretasi <i>Cluster</i> Optimal	64
4.7. Pemetaan Daerah Terdampak Covid 19 di Jawa Timur Bulan April-Desember 2020	66
4.8. Integrasi Keislaman	74
V PENUTUP	77

s Covid-19 di Jawa Timur Bulan Juli 2020	..
s Covid-19 di Jawa Timur Bulan Agustus 2020	
s Covid-19 di Jawa Timur Bulan September 2020	
s Covid-19 di Jawa Timur Bulan Oktober 2020	
s Covid-19 di Jawa Timur Bulan Nopember 2020	
s Covid-19 di Jawa Timur Bulan Desember 2020	
TANDARISASI DATA <i>Z-SCORE</i>
RAK EUCLIDEAN
uster dan Anggota Metode <i>Single Linkage</i> Bulan
uster dan Anggota Metode <i>Complete Linkage</i> Bulan
uster dan Anggota Metode <i>Average Linkage</i> Bulan
uster dan Anggota Metode <i>HK-Means Single Linkage</i> Bulan
2020

DAFTAR TABEL

2.1	Nilai <i>Kaufman</i>	26
3.1	Variabel penelitian yang digunakan	33
4.1	Perhitungan Uji Multikolineritas Variabel Konfirmasi X_1 dan Sem- buh X_2	40
4.2	Perhitungan Uji Multikolineritas Pada Semua Variabel	42
4.3	Anggota tiap <i>cluster</i> dari pembentukan 2 <i>cluster Single Linkage</i>	55
4.4	Pusat <i>cluster (centroid)</i> pada 2 <i>cluster single linkage</i>	56
4.5	Jarak <i>Eulidean</i> dan Hasil <i>Cluster</i> Iterasi ke-1	57
4.6	Pusat <i>cluster (centroid)</i> pada 2 <i>cluster single linkage</i>	60
4.7	Nilai <i>Sillhoutte Coefficient</i> masing-masing percobaan <i>cluster</i>	61
4.8	<i>centroid</i> Akhir pada 2 <i>cluster Complete linkage</i>	65
4.9	<i>Centroid</i> Akhir pada 2 <i>cluster Complete linkage</i>	65
4.11	Kabupaten atau kota berdasarkan hasil analisis <i>cluster</i> menggunak- an <i>HK-Means(Complete Linkage)</i>	66
4.12	Pemetaan Kabupaten/Kota Berdasarkan Hasil Analisis <i>Cluster</i>	67
5.1	Nilai pusat <i>cluster</i> pada 2 <i>cluster Complete linkage</i>	77

Melindungi keselamatan diri ialah kewajiban tiap muslim. Dalam bahasa ushul fiqih, diketahui dengan sebutan *hifz an-nafs* (melindungi jiwa, melindungi keselamatan diri). Melindungi keselamatan diri ialah bagian utama dalam agama Islam serta ialah salah satu tujuan utama dalam hukum Islam. Imam Al - Ghazali, berkata mewujudkan *hifz an-nafs* sama dengan melindungi agama Islam. Dalam konteks pandemi Covid - 19 dikala ini, realisasi melindungi keselamatan diri terwujud dalam wujud beribadah di rumah serta melaksanakan protokol kesehatan. Perihal ini ialah wujud usaha nyata buat menghindari penularan virus tersebut. Tidak tertular virus ini sama dengan melindungi kesehatan serta keselamatan nyawa kita. Dalam Alqur'an surat Al-Maidah ayat 32 ditegaskan bahwa menyelamatkan satu nyawa sama dengan menyelamatkan seluruh nyawa manusia :

مِنْ أَجْلِ ذَلِكَ كَتَبْنَا عَلَىٰ بَنِي إِسْرَآئِيلَ أَنَّهُ مَنْ قَتَلَ نَفْسًا بِغَيْرِ نَفْسٍ أَوْ
فَسَادٍ فِي الْأَرْضِ فَكَأَنَّمَا قَتَلَ النَّاسَ جَمِيعًا وَمَنْ أَحْيَاهَا فَكَأَنَّمَا أَحْيَا
النَّاسَ جَمِيعًا وَلَقَدْ جَاءَتْهُمْ رُسُلُنَا بِالْبَيِّنَاتِ ثُمَّ إِنَّ كَثِيرًا مِنْهُمْ بَعْدَ ذَلِكَ
فِي الْأَرْضِ لَمُسْرِفُونَ

[illegible]

Pada 12 Maret 2020 WHO (*World Health Organization*) mengumumkan bahwa Covid-19 sebagai pandemik. Virus ini sudah menyebar di 215 negara di dunia termasuk Indonesia. Berdasarkan informasi yang dipublikasikan oleh Kementerian Kesehatan Republik Indonesia sejak pertama kali dilaporkan pada bulan Desember 2019 sampai tanggal 31 Desember 2020, di Indonesia telah terkonfirmasi kasus Covid-19 sebanyak 743.198 kasus dengan jumlah kematian sebanyak 22.138 kasus.

[illegible]

Tingginya penyebaran Covid-19 di Jawa Timur ini maka perlu adanya pengelompokan daerah terdampak Covid-19 di Jawa Timur untuk mengetahui kemiripan karakteristik atau kriteria dari setiap daerah. Daerah yang memiliki karakteristik yang sama atau hampir sama akan terkumpul dalam suatu kelompok. Hal ini bertujuan untuk memberikan informasi terkait Covid-19 kepada masing-masing daerah. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu dengan analisis *cluster*.

Pada penelitian sebelumnya yaitu penelitian yang dilakukan oleh Rini Silvi (2018) yang membandingkan metode *centroid linkage* dan *K-means* dalam menentukan hasil analisis *cluster* dengan data outlier untuk pengelompokan indikator HIV/AIDS di Indonesia. Penelitian ini menggunakan gap statistik untuk menentukan jumlah *cluster* ideal yang mengelompokkan propinsi berdasarkan indikator HIV/AIDS sedemikian hingga terbagi menjadi 7 *cluster*. Hasil penelitian ini memberikan kesimpulan bahwa untuk data yang memiliki *outlier*, metode pengclusteran menggunakan *centroid linkage* lebih memberikan hasil yang sesuai dibandingkan dengan metode *K-means*. Rasio Sw/Sb metode *K-means* adalah sebesar 0,112232 dan rasio Sw/Sb metode *centroid linkage* menghasilkan angka 0,067307. Metode *centroid linkage* menghasilkan kelompok yang lebih homogen sehingga nilai rasio yang dihasilkan lebih kecil. Artinya metode *centroid linkage* memiliki kualitas ketepatan kelompok yang lebih baik dibandingkan metode *K-means*. (Silvi, 2018)

Selain itu penelitian lainnya juga dilakukan oleh Lynda Rahmawati, Sari Wi-
dya Sihwi dan Esti Suryani (2016) dengan judul Analisa *clustering* menggunakan
metode *K-means* dan *hierarchical clustering* kasus : dokumen skripsi jurusan ki-
mia, FMIPA Universitas Sebelas Maret. Pengelompokkan data penelitian yang ber-

Analisa Perbandingan Metode *Hierarchical Clustering*, *K-means* dan Gabungan Keduanya dalam *Cluster Data* (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS) yang dilakukan oleh Tahta Alfina, Budi Santosa, dan Ali Ridho Barakbah. Clustering memiliki dua metode, yaitu partisi dan hierarki. Dua metode ini memiliki kelebihan dan kekurangan masing-masing, dan dengan menggabungkan keduanya dapat diperoleh hasil cluster yang lebih baik. Dari hasil *cluster* dengan menggunakan data problem Kerja Praktek Jurusan Teknik Industri ITS, maka diperoleh hasil bahwa kombinasi algoritma *hierarchical clustering* dan *k-means* menghasilkan pengelompokan data yang lebih baik. Pada skenario jumlah *cluster* sebanyak 5, penjumlahan nilai *s* terbesar diperoleh oleh metode *single linkage clustering* yang dikombinasikan dengan *k-means* sebesar 300 data diclusterkan pada *cluster* yang tepat dengan nilai koefisien kolerasi sebesar 0,9118, diikuti oleh 3 metode *hierarchical clustering* yang lainnya yang digabungkan dengan *k-means* dan penjumlahan nilai *s* yang paling kecil dihasilkan oleh metode *k-means* yaitu sebesar 210 data. (Alfina et.all , 2019)

[illegible]

19 menghasilkan sejumlah 33 kabupaten/kota berada pada cluster 1, 1 kota berada pada cluster 2, dan 1 kabupaten berada pada clusters 3. Kondisi kasus COVID-19 di Provinsi Jawa Tengah sudah relatif aman tetapi harus tetap dilakukan penanganan semaksimal mungkin agar kasus aktif dan kasus meninggal tidak semakin banyak. Selain itu, cluster 2 yaitu Kota Semarang dan cluster 3 yaitu Kabupaten Jepara harus lebih diperhatikan oleh Pemerintah Provinsi Jawa Tengah karena wilayah tersebut masih memiliki jumlah kasus aktif yang tinggi. Hal tersebut harus dilakukan agar tidak terjadi pertambahan kasus meninggal maupun kasus aktif di wilayah tersebut atau penyebaran kasus aktif di luar wilayah tersebut. (Mahmudah, 2020)

Berdasarkan pada penjelasan tersebut, penulis tertarik dalam melakukan penelitian mengenai analisis cluster dengan mengambil kasus Covid-19 yang sekarang menjadi pandemi di seluruh dunia. Judul dalam penelitian ini adalah " Pengelompokan Daerah Terdampak Covid-19 Di Provinsi Jawa Timur Menggunakan Metode *Hierarchical K-Means Clustering (HK-Means)* (Studi Kasus : Jawa Timur)". Metode yang digunakan tersebut merupakan gabungan antara metode *hierarchical clustering* dan *K-means*, digabungkan agar dapat meningkatkan dan menghasilkan hasil *cluster* dengan kualitas yang lebih baik dan tinggi dan pada berdasarkan penelitian alfini terbukti bahwa gabungan metode tersebut menghasilkan hasil yang lebih baik dibandingkan dengan metode *K-means*. Karena hasil akhir pengelompokan dengan metode *K-means* sangat sensitif terhadap pemilihan pusat *cluster* jadi hasilnya sedikit berbeda setiap kali dihitung. Metode *K-means* mempunyai kelemahan yaitu pada penentuan pusat awal *cluster*. Hasil *cluster* yang terbentuk dari metode *K-Means* ini sangatlah tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan. Untuk itu, metode *K-means* digabungkan dengan metode *Hierarchical Clustering* untuk penentuan pusat awal *cluster* untuk menghasilkan nilai yang le-

Pada penelitian ini dimulai dengan tahapan Hierarchical clustering terlebih dahulu untuk menghasilkan *centroid* (pusat *cluster*) awal menggunakan jarak *Euclidean* pada metode *Singele Linkage*, *Complete Linkage*, *Average Linkage*. Dari hasil *Hierarchical Clustering* digunakan sebagai *centroid* (pusat *cluster*) awal untuk tahapan *K-Means Clustering*. Penelitian ini dilakukan dengan menggunakan gabungan metode *Hierachical clustering* yaitu metode (*Single Linkage*, *Complete Linkage*, *Average Linkage*) dengan *K-means* sehingga dapat diketahui gabungan metode manakah yang menghasilkan hasil *cluster* yang terbaik. Pengelompokan yang memiliki kinerja terbaik dilihat dari nilai *Sillhoutte Coefficient*.

Menurut latar belakang masalah yang sudah dijelaskan, bahwa permasalahan pada penelitian ini:

- ### 1.3. Tujuan Penelitian

1. Untuk menentukan metode gabungan *HK-Means* apakah yang menghasilkan hasil *cluster* terbaik dilihat dari nilai *Silhouette Coefficient* pada data Covid-

- #### 1.4. Manfaat Penelitian

- Penelitian ini dapat menambah ilmu pengetahuan dan wawasan mengenai analisis *cluster* dan gabungan metode yang terbaik dalam pengelompokan daerah terdampak Covid-19 di Jawa Timur.

- Bagi masyarakat umum, penelitian ini dapat memberikan informasi tentang sebaran kabupaten/kota terdampak Covid-19 di Provinsi Jawa Timur. Sehingga dapat melakukan pencegahan agar tidak semakin menyebar luas.

1. Data yang digunakan pada penelitian ini merupakan data penyebaran Covid-19 di Propinsi Jawa Timur pada bulan April-Desember 2020 yang diakses melalui laman resmi pemerintah Propinsi Jawa Timur.
2. Penelitian ini menggunakan jarak *Euclidean* dalam perhitungan kemiripan antar obyek atau daerah.
3. Menggunakan metode *Single Linkage*, *Complete Linkage* dan *Average Linkage* dalam melakukan analisis *cluster*.

Sistematika penulisan yang digunakan pada penyusunan skripsi ini adalah sebagai berikut:

Bab ini memaparkan tentang latar belakang penulisan skripsi, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

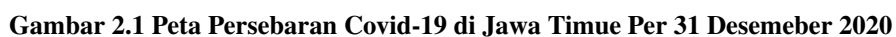
Bab ini menjelaskan tentang penjelasan tema yang dijadikan landasan teori seperti Covid-19, analisis *cluster*, standarisasi data, penetapan metode analisis cluster menggunakan metode *HK-Means*, pemilihan metode terbaik menggunakan *Sillhoutte Coefficient*.

Bab ini menjelaskan tentang jenis-jenis penelitian, cara pengumpulan data, variabel penelitian, dan cara analisis data.

Bab ini menjelaskan hasil dari penelitian tentang pengelompokan daerah terdampak Covid-19 di Jawa Timur menggunakan metode *Hierarchical K-Means Clustering* (HK-Means).

Bab ini menjelaskan tentang kesimpulan dari penelitian, beserta saran yang ditujukan kepada pembaca dan peneliti selanjutnya.

Khususnya di provinsi Jawa Timur yang merupakan salah satu wilayah di Indonesia yang penyebaran Covid-19 juga terus meningkat dan belum menunjukkan penurunan sampai saat ini. Pada tanggal 31 Desember 2020 melalui laman web covid Jawa Timur bahwa jumlah kasus positif Covid-19 di Jawa Timur mengalami penambahan kasus baru sebanyak 935 kasus baru jadi total sebanyak 84.152 kasus dan jumlah kasus yang sembuh sebanyak 72.135 kasus dengan jumlah kematian sebanyak 5.827 kasus. Dengan penambahan kasus tersebut membuat Jawa Timur menduduki peringkat empat kasus Covid-19 terbanyak setelah DKI Jakarta dan Jawa Tengah. Bahkan Jawa timur sempat menempati peringkat pertama kasus covid-19 terbanyak yaitu pada 28 Juni 2020 dengan penambahan kasus baru sebanyak 330 kasus. (Satgas COVID-19 JATIM, 2020)



Standarisasi data digunakan apabila pada variabel data yang digunakan terdapat perbedaan ukuran atau nilai skala yang besar. Standarisasi data yang sering digunakan yaitu standarisasi data dengan *z-score* dengan mengkonversi nilai tiap variabel dengan nilai standar (*z-score*) yakni dengan menghitung nilai tengah dan membagi hasilnya dengan standar deviasi dari masing variabel. Berikut merupakan rumus *z-score* : (Walpole at.all , 1995)

Keterangan :

s = standar deviasi

Terdapat angapan-anggapan yang perlu diamati saat melakukan analisis *cluster*, salah satunya ialah uji Multikolinieritas. Multikolinearitas adalah adanya hu-

Dimana :

Yang dirumuskan dengan:

$$R_j^2 = \left[\frac{n \sum_{i=1}^n X_1 X_2 - \sum_{i=1}^n X_1 \sum_{i=1}^n X_2}{\sqrt{\left(n \sum_{i=1}^n (X_1)^2 - \left(\sum_{i=1}^v X_1 \right)^2 \right) - \left(n \sum_{i=1}^n (X_2)^2 - \left(\sum_{i=1}^v X_2 \right)^2 \right)}} \right]^2 \quad (2.3)$$

$$Tolerance = \frac{1}{VIF} \quad (2.4)$$

Cara untuk menanggulangi timbulnya multikolinieritas yaitu dengan mengecualikan variabel yang tidak sesuai atau relevan. Dengan cara memilih salah satu

2.4. Jarak *Euclidean*

$$d - ij = \sqrt{\sum_{k=1}^p x_i k x_i j^2} \quad (2.5)$$

$d_{i,j}$ = jarak *Euclidean* dari obyek ke- i dan obyek ke- j

x_{ik} = nilai objek ke- i pada variabel ke- k

x_{jk} = nilai objek ke- j pada variabel ke- k

Clustering atau analisis pengelompokkan merupakan suatu proses membagi data dari suatu himpunan tertentu kedalam beberapa kelompok yang mempunyai kesamaan karakteristik data dengan karakteristik data dalam kelompok lain. Kelas

Larose dalam buku yang ditulis oleh Kusrini dan Luthfi mengelompokkan Data Mining dapat dibagi menjadi 6 kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi, *clustering* (pengelompokan), dan asosiasi. Berbeda dengan klasifikasi, teknik *clustering* yaitu mengelompokkan data secara otomatis sebelum mengetahui label kelasnya. Tetapi *clustering* juga dapat digunakan untuk himpunan data dengan label kelas yang diketahui (Suryanto , 2017).

Sehingga, dari semua pengertian bahwa *clustering* adalah suatu metode yang digunakan untuk memberikan label bagi data yang belum diketahui secara pasti kelas *cluster*. Ciri-ciri *cluster* adalah sebagai berikut (Santoso dan Slamet , 1987) :

- Menurut (N. Alshatri et.al , 2014), *clustering* dikelompokkan menjadi 5 kelompok besar yaitu :

Algoritma *Partitional Clustering* membagi titik data ke dalam partisi k , di mana setiap partisi mewakili sebuah *cluster*. Contoh algoritma *partitio-*

ning based clustering adalah *k-means*, *k-medoids*, *k-modes*, *PAM*, *CLARANS*, *FCM* (Chitra dan Maheswari, 2017).

2. Hierarchical Based Clustering

Hierarchical Based clustering adalah metode analisis *cluster* yang digunakan untuk membangun hierarki *cluster*. Dalam data mining, cara kerja *hierarchical clustering* ialah dengan mengelompokkan objek-objek data ke dalam pohon *cluster*, yang mana struktur pohon *cluster* ini biasa disebut dendrogram. Contoh algoritma *hierarchical based clustering* adalah *BIRCH*, *CURE*, *ROCK*, dll (Rani dan Rohil at.all , 2013).

3. Density Based Clustering

Algoritma ini dapat menemukan *cluster* yang berbentuk acak dan dapat menangani *noise*, dan juga merupakan algoritma satu-scan yang mana hanya melakukan satu kali *scanning* ketika memeriksa data mentah. Contoh algoritma *density based clustering* ialah *DBSCAN*, *OPTICS*, *LDBSCAN*, dll (Amini et.al , 2011).

4. Grid Based Clustering

Algoritma *Grid-Based Clustering* adalah tipe yang paling penting dalam algoritma *clustering* hierarki. Pendekatan *Grid-Based* lebih menganggap sel dibandingkan titik data. Karena sifat algoritma *Grid-Based Clustering* ini umumnya lebih efisien secara komputasi apabila dibandingkan dengan semua jenis algoritma *clustering*. Contoh algoritma *grid based clustering* adalah *wave-cluster*, *STING*, *CLIQUE*, *OptiGrid* (N. A. Shah dan Paul , 2017).

5. Model Based Clustering

Metode *Hierarchical Clustering* adalah metode yang memulai pengelompokannya dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses dilanjutkan ke obyek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam “pohon” dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai dengan yang paling tidak mirip. Biasanya pengelompokan ini disajikan dalam bentuk dendrogram, yang mirip dengan “struktur diagram pohon” (*tree diagram*).

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk
2. Setiap data obyek dianggap sebagai *cluster* sehingga $n = N$
3. Menghitung jarak antar *cluster*.
4. Mencari dua *cluster* yang mempunyai jarak antar *cluster* paling minimal dan menggabungkannya (berarti $N = n - 1$) jika $n > k$, maka kembali ke langkah 3.

[illegible]

1. Single Linkage

Single Linkage adalah proses pengklasteran yang didasarkan pada jarak terdekat antar obyeknya (minimum distance). Metode ini sangat bagus untuk melakukan analisa pada tiap tahap pembentukan *cluster*. Metode ini juga sangat cocok untuk dipakai pada kasus *shape independent clustering* karena kemampuannya untuk membentuk pola tertentu dari *cluster*. Metode ini dikenal pula dengan pendekatan tetangga terdekat (*nearest neighbor*) dalam kelompok yang berbeda. Berbeda dengan prosedur *Complete Linkage*, pada prosedur ini pengelompokkan dilakukan berdasarkan jarak minimum. Jika individu X dan Y mempunyai jarak d_{xy} terdekat, maka harus dicari jarak minimum XZ dan XY , sehingga:

$$d_{XY}Z = Min(d_{XZ}, d_{YZ}) \quad (2.6)$$

[illegible]

tersebut merupakan kelompoknya.

2. Complete Linkage

Metode *Completed linkage* adalah metode pautan lengkap yang dikembangkan oleh Horn (1943) dan Sorensen (1948). Metode ini sering dikenal dengan metode maksimum (Latan, 2014). Metode *Completed linkage* berdasar pada jarak maksimum. Pengukuran jarak antara *cluster* satu dengan yang lainnya menurut pada objek yang mempunyai jarak paling jauh. Metode *Complete Linkage* ditentukan dengan memakai jarak terjauh antara dua objek pada kelompok yang berbeda (*Furthest Neighbor*). Dalam perhitungan awal, nilai minimum dalam $D = d_{ij}$ dicari dahulu kemudian digabungkan pada objek-objek yang bersesuaian, contohnya I dan K , untuk menghasilkan *cluster* (IJ). Jarak antara *cluster* lain K dan (IJ), dihitung dengan cara :

$$d_{IJ}K = \text{Max}(d_{IK}, d_{JK}) \quad (2.7)$$

d_{IK} dan d_{IK} adalah jarak terjauh diantara anggota *cluster-cluster* I dan K juga *cluster-cluster* J dan K .

3. Average Linkage

Metode *Average Linkage* adalah proses *clustering* (*Average Distance*) atau pada jarak rata-rata antar obyeknya. Untuk metode *Average Linkage*, jarak diantara dua *cluster* dapat dimisalkan sebagai jarak rata-rata dari semua anggota dalam satu *cluster* dengan semua anggota dalam *cluster* lainnya. Metode ini mempunyai tujuan untuk meminimalkan rata-rata jarak semua pasangan dan dua kelompok yang digabungkan. Langkah awal yang perlu dilakukan pada metode ini adalah sama dengan metode-metode sebelumnya.

Keterangan :

$$d_{IJ}K = \frac{\sigma_a \sigma_b d_{ab}}{N_{IJ} N_K} \quad (2.8)$$

$$N_{IJ} = \text{jumlah obyek pada } cluster(IJ)$$

2.6.2. Metode *Divisive* (pemecahan)

Untuk metode *Divisive*, pemecahan diawali dari satu *cluster* besar yang berisi seluruh observasi. Kemudian untuk observasi yang berbeda akan dipisahkan dan dibentuk kelompok-kelompok yang lebih kecil. Proses ini akan terus dikerjakan sampai setiap observasi terbentuk kelompok masing-masing.

2.7. K-Means

Salah satu metode *Non-Hierarchical Clustering* yang terkenal dan sederhana adalah metode *K-means*. *K-means* ditemukan oleh banyak orang yaitu seperti Lloyd (1957, 1982), Forgey (1965), Friedman dan Rubin (1967), McQueen (1967). Pada tahun 1957 ide dari *clustering* pertama kali ditemukan oleh Lloyd, namun hal tersebut baru diterbitkan pada tahun 1982 (Yulian, 2018). Algoritma *K-means* adalah suatu algoritma pengelompokan yang melakukan partisi set data ke dalam sejumlah *K-cluster* yang sudah ditentukan di awal. *K-means* akan mempartisi data penelitian ke dalam bentuk beberapa *cluster* atau kelompok sehingga data yang memiliki karakteristik sama akan dikelompokkan menjadi satu *cluster* yang sama dan

$d(x_i, x_j)$ = Jarak antara objek ke- i dengan objek ke- j , dimana $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, n$.

x_{jl} = Nilai objek ke-j pada variabel k.

n = Banyaknya variabel data

1. Menentukan *k-cluster* yang digunakan.
2. Menentukan pusat awal *cluster(centroid)* secara acak dari objek data sebanyak *k*.
3. Menentukan jarak terdekat dari setiap objek pengamatan dengan *centroid* yang telah ditentukan menggunakan jarak *euclidean* dengan persamaan :

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^n (x_{il} - x_{jl})^2} \quad (2.9)$$

Keterangan :

$d(x_i, x_j)$ = Jarak antara objek ke-*i* dengan objek ke-*j*, dimana $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, n$.

x_{il} = Nilai objek ke-*i* pada variabel *k*

x_{jl} = Nilai objek ke-*j* pada variabel *k*.

n = Banyaknya variabel data
4. Menentukan jarak terdekat suatu objek dengan pusat *cluster centroid*.
5. Menentukan pusat *cluster (centroid)* baru dengan menghitung rata-rata masing-masing *cluster* menggunakan :

$$ckl = \frac{x_{1l} + x_{2l} + \dots + x_{pl}}{p} \quad (2.10)$$

[illegible]

Silhouette coefficient merupakan metode evaluasi untuk menguji optimal atau ketepatan sebuah *cluster* yang telah terbentuk dari proses clustering. *Silhouette coefficient* menekankan pada validasi dan penafsiran set data yang dikembangkan. Penggabungan konsep *cohesion* dan *separation* digunakan untuk memvalidasi kekuatan struktur hasil clustering. Tahapan perhitungan *silhouette coefficient* adalah sebagai berikut (Handoyo et.al , 2014)

- $$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq id(i, j)} \quad (2.12)$$

$A = Cluster$

- $$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.13)$$

$C = \text{Cluster lain selain cluster A atau cluster C tidak sama dengan cluster.}$

3. Setelah menghitung $d(i, C)$ untuk semua C , maka diambil nilai terkecil dengan menggunakan Persamaan 2.14.

$$b(i) = \min C \neq Ad(i, C) \quad (2.14)$$

Keterangan :

$$b(i) = \text{Tetangga dari objek } i$$

4. Menghitung nilai *silhouette coefficient* dengan Persamaan 2.15.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.15)$$

Keterangan :

$s(i)$ = Nilai *silhouette coefficient* data ke i $a(i)$ = Rata-rata jarak suatu data ke

$i \ b(i)$ = Rata-rata jarak suatu data ke semua objek

Hasil perhitungan nilai *silhouette coefficient* bervariasi dengan rentang -1 sampai 1. Apabila nilai *silhouette coefficient* bernilai positif yaitu ($a(i) < b(i)$) dan $a(i)$ mendekati 0 maka nilai *clustering* nya baik, sehingga nilai *silhouette coefficient* yang dihasilkan maksimal yaitu 1. Objek i dikatakan berada pada *cluster* yang tepat apabila $SI = 1$. Sedangkan, apabila $SI = 0$ maka objek i berada diantara dua cluster artinya struktur yang dimiliki objek tersebut tidak jelas. Lebih lanjut, apabila $SI = 1$ maka objek tersebut memiliki nilai *overlapping* yang berakibat dimasukkan dalam *cluster* lain.

Nilai rata-rata yang dimiliki oleh *silhouette coefficient* dari masing-masing data objek dalam suatu *cluster* menunjukkan seberapa layak data tersebut dimasukkan dalam *cluster*. Berikut adalah tabel nilai Kaufman berdasarkan (Rousseeuw ,

Keterangan :

2.10. Kewajiban Menjaga Kesehatan dan Keselamatan Jiwa

Dalam islam terdapat lima kebutuhan penting yang harus dijaga oleh umat islam (*Dharuriyyatul-Khams*). Lima kebutuhan penting yang harus dijaga meliputi, meliputi penjagaan terhadap *din* (agama), keselamatan jiwa, keturunan, akal, dan harta. Maka, menjaga kesehatan dan keselamatan diri merupakan kewajiban setiap Muslim. Menjaga keselamatan diri dalam bahasa ushul fiqh, dikenal dengan istilah *hif an-nafs* (menjaga jiwa, menjaga keselamatan diri). Menjaga keselamatan diri sama dengan menjaga agama islam. Oleh karena itu Rasulullah SAW melarang berpuasa bagi orang yang sakit dan sedang dalam keadaan musafir(dalam perjalanan

Untuk menghitung varian *Silhouette Coefficient* menggunakan Persamaan 2.16

$$s^2 = \frac{n \sum_{i=1}^n s(i)^2 - \sum_{i=1}^n (s(i))^2}{n(n-1)} \quad (2.16)$$

s^2 = Nilai varian *silhouette coefficient*
 $s(i)$ = Nilai *silhouette coefficient* data ke i
 n = Banyaknya data

2.10. Kewajiban Menjaga Kesehatan dan Keselamatan Jiwa

hu 'alaihi wasallam bersabda: "Tutuplah bejana-bejana
 minuman, karena di suatu malam pada setiap tahunnya
 (perbahaya) yang akan jatuh ke dalam bejana dan ket
 tutup." Dan telah menceritakan kepada kami Nashr b
 ceritakan kepadaku Bapakku; Telah menceritakan k
 n Hadits dan sanad yang serupa, hanya saja dia ber
 suatu hari pada setiap tahunnya akan ada wabah pe
 pada akhir Haditsnya; Al laits berkata; 'Orang-or
 diantara kami merasa takut pada hal itu sejak bulan

Mencuci tangan

asa mencuci tangan. Mencuci tangan ketika pandemi

diri dari paparan Covid-19, karena penularan V

قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ الطَّاعُونَ آيَةُ الرَّجْرِ ابْتُلَى اللَّهُ عَزَّ وَجَلَّ بِهِ نَاسًا مِنْ عِبَادِهِ فَإِذَا سَمِعْتُمْ بِهِ قَوْلًا تَدْخُلُوا عَلَيْهِ وَإِذَا وَقَعَ بِأَرْضٍ وَأَنْتُمْ بِهَا فَلَا تَمُرُّوا مِنْهُ

Dalam situasi dimana masyarakat dihadapkan kepada suatu wabah seperti Covid-19 ini, maka mereka harus (wajib) untuk menjauhkan diri atau menghindarkan diri dari wabah itu agar tidak terkena wabah. Sebaliknya, apabila individu sedang berada di sebuah wilayah atau lingkungan yang tengah terkena pandemi Covid-19 atau wabah lainnya, maka dia dilarang (haram) untuk keluar dari lingkungan atau wilayah itu agar wabah yang berada pada lingkungannya tidak meluas ke daerah lain. (Arifin et.all, 2019).

Sebagai seorang muslim kita tidak boleh berprasangka buruk Ta'ala, khususnya ketika sedang menghadapi bala' dan bencana. Allah Ta'ala berfirman:

إِذْ جَاءُوكُم مِّن فَوْقِكُمْ وَمِنْ أَسْفَلَ مِنكُمْ وَإِذْ زَاغَتِ الْأَبْصَارُ وَبَلَغَتِ الْقُلُوبُ الْحَنَاجِرَ وَتَظُنُّونَ بِاللَّهِ الظُّنُونَا

[illegible]

2.10.5. Bersikap Optimis dan Berucap yang Baik

Hal ini sebagaimana diajarkan oleh Nabi dalam hadits dari Anas bin Malik RA. Yang berarti "Tidaklah penyakit menular tanpa izin Allah dan tidak ada pengaruh dikarenakan seekor burung, tetapi yang mengagumkanku ialah al-Fa'lu (optimisme), yaitu kalimah hasanah atau kalimat thayyibah (kata-kata yang baik)." (HR. Al-Bukhari, Muslim).

Tabel 3.1 Variabel penelitian yang digunakan

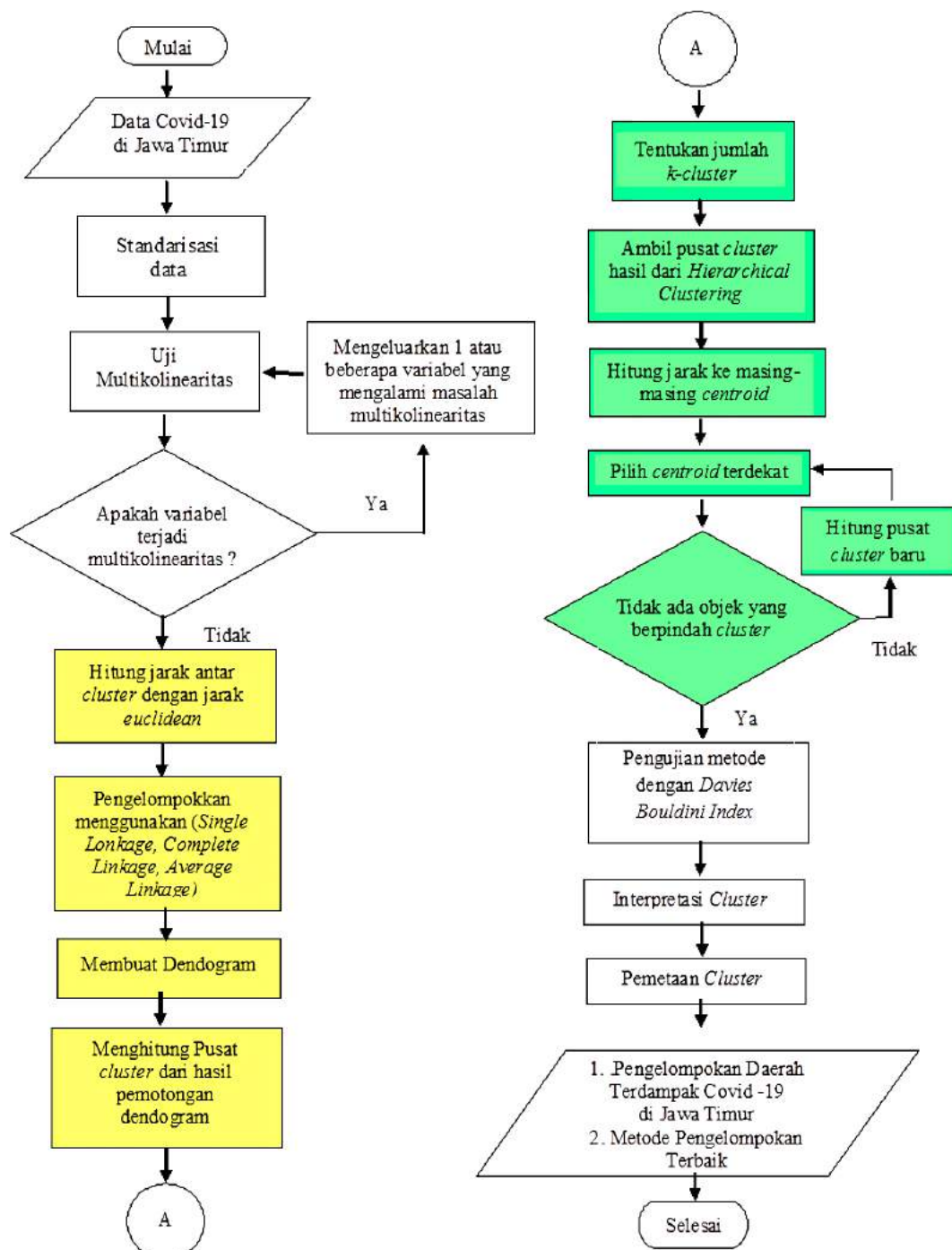
Variabel	Keterangan
x_1	Konfirmasi yaitu jumlah kasus yang terkonfirmasi Covid-19 pada setiap kota/kabupaten
x_2	Sembuh yaitu jumlah kasus yang sembuh dari Covid-19 pada setiap kota/kabupaten
x_3	Meninggal yaitu jumlah kasus yang meninggal karena Covid-19 pada setiap kota/kabupaten
x_4	Bergejala yaitu jumlah kasus yang berasal dari kasus suspek (bergejala)dari Covid-19 pada setiap kota/kabupaten
x_5	Tanpa Gejala yaitu jumlah kasus bukan berasal dari kasus suspek (bergejala)dari Covid-19 pada setiap kota/kabupaten
x_6	Perjalanan yaitu jumlah kasus Covid-19 dengan risiko riwayat perjalanan/pelaku perjalanan pada setiap kota/kabupaten
x_7	Kontak yaitu jumlah kasus Covid-19 yang konfirmasi dengan risiko riwayat kontak erat pada setiap kota/kabupaten
x_8	Tanpa riwayat yaitu jumlah kasus Covid-19 yang terkonfirmasi tanpa riwayat pada setiap kota/kabupaten

Data diperoleh dari laman resmi pemerintah provinsi Jawa Timur.

(Satgas COVID-19 JATIM , 2020)

3.4. Tahapan Penelitian

Pada tahap ini dilakukan pengelompokan data menggunakan metode *HK-Means*, yaitu gabungan *Hierarchical clustering* dan metode *K-means*. Gambar 3.1 adalah flowchart yang menjelaskan urutan pengerjaan penelitian dengan menggunakan metode *HK-Means*, yang berwarna hijau untuk proses *Hierarchical clustering* dan berwarna kuning untuk proses *K-means*.



Gambar 3.1 Diagram Alir Metode

1. Memasukkan data Covid-19 di Jawa Timur.
2. Melakukan standarisasi data dengan menggunakan Persamaan 2.1

3. Melakukan uji multikolinearitas terhadap data untuk mengetahui suatu data terjadi korelasi antar variabel atau tidak dengan menggunakan persamaan 2.2
4. Menentukan nilai k . Pada penelitian ini dilakukan ujicoba k dengan $k = 2, 3, 4$, dan 5.
5. Melakukan pengelompokkan dengan metode *Hierarchical clustering* menggunakan teknik pengelompokkan *Agglomeratif* yaitu *Single Linkage*, *Complete Linkage* dan *Average Linkage*.
6. Melakukan pengelompokan setiap daerah dengan metode *Single Linkage*. Tahapan pengelompokan dengan metode *Single Linkage*, *Complete Linkage*, dan *Complete Linkage* sebagai berikut :
 - (a) Menghitung jarak antar objeknya dengan jarak *euclidean* dengan Persamaan 2.5.
 - (b) Kemudian, dari hasil perhitungan tadi dipilih jarak yang paling minimal atau yang terdekat antar objek.
 - (c) Merhitungan ulang jarak dengan metode *Single Linkage* dengan Persamaan 2.6, metode *Complete Linkage* dengan Persamaan 2.7 dan metode *Average Linkage* dengan Persamaan 2.8.
 - (d) Hal ini akan terus dilakukan dan akan berhenti jika memenuhi kondisi jumlah $k = 1$ atau tersisasa satu *cluster*.
 - (e) Menentukan jumlah anggota *cluster* beserta anggotanya dengan membentuk dendrogram yang sesuai dengan *cluster*.
7. Pada akhir tahap *Hierarchical Clustering* ini akan diperoleh sebuah gambar dendrogram yang menunjukkan urutan pengelompokan masing-masing anggota dalam *cluster*.

(a) Menentukan pusat *cluster* pusat awal (*centroid*) untuk metode *K-means* dengan mencari rata-rata dari data yang berada pada sebuah *cluster* hasil dari *Hierarchical Clustering*.

(c) Menentukan jarak terdekat dari setiap objek pengamatan dengan *centroid* yang telah ditentukan menggunakan jarak *euclidean* dengan Persamaan 2.9.

(e) Menentukan pusat *cluster* (*centroid*) yang baru dengan menggunakan Persamaan 2.10.

(f) Melakukan iterasi pada *K-Means* dan iterasi akan berhenti apabila semua data yang berada pada suatu *cluster* tertentu dan tidak berpindah ke *cluster* yang lainnya dan memiliki nilai pusat *cluster* (*centroid*) baru dan nilai pusat *cluster* (*centroid*) lama sama.

10. Melakukan pengujian gabungan metode menggunakan *Sillhoutte Coefficient* dan dipilih hasil yang terbaik.

11. Melakukan interpretasi *cluster*. Misalnya terbentuk 3 *cluster*, maka perlu ditentukan mana *cluster* yang tinggi, sedang dan rendah.

12. Melakukan pemetaan *cluster*.

Gambar 4.2 merupakan diagram data kasus Covid-19 setiap kabupaten atau kota di Jawa Timur pada Desember 2020. Dapat dilihat bahwa Kabupaten Jember memiliki jumlah konfirmasi dan sembuh tertinggi disusul Kabupaten Banyuwangi dan Kota Surabaya.

--	--	--	--	--	--

Kabupaten/Kota	X_1	X_2	$X_1 * X_2$	$(X_1)^2$	$(X_2)^2$
Kab.Situbondo	0,1061	0,4833	0,0513	0,0113	0,2336
Kab.Madiun	-0,9224	-0,9778	0,9020	0,8509	0,9562
Kab.Sumenep	-0,2831	-0,7039	0,1992	0,0802	0,4954
Kota Surabaya	1,6040	1,9141	3,0703	2,5729	3,6637
Kab.Gresik	-0,5709	-0,4991	0,2849	0,3260	0,2491
Kab.Pamekasan	-0,5709	-0,7841	0,4477	0,3260	0,6148
Kab.Bojonegoro	0,0188	-0,1864	-0,0035	0,0004	0,0347
Kota Mojokerto	-0,8328	-0,7232	0,6023	0,6935	0,5231
Kab.Lamongan	0,2830	0,4031	0,1141	0,0801	0,1625
Kab.Trenggalek	-0,4954	-0,4105	0,2034	0,2455	0,1685
Kab.Pasuruan	-0,5898	-0,6015	0,3548	0,3479	0,3618
Kota Blitar	-0,4058	-0,2555	0,1037	0,1647	0,0653
Kab.Pacitan	-0,4412	-0,4603	0,2031	0,1947	0,2120
Kab.Banyuwangi	2,2009	2,0303	4,4684	4,8437	4,1222
Kab.Kediri	1,4059	1,2417	1,7456	1,9765	1,5416
Kota Kediri	-0,6818	-0,5987	0,4082	0,4649	0,3584
Kab. Ponorogo	-0,0166	-0,0535	0,0009	0,0003	0,0029
Kab. Blitar	0,2878	0,3754	0,1080	0,0828	0,1409
Kab.Sampang	-0,9224	-0,8782	0,8101	0,8509	0,7712
Kota Malang	1,9720	1,4132	2,7869	3,8889	1,9971
Kab.Jember	3,3284	3,6880	12,2752	11,0785	13,6012

$$R_j^2 = \left[\frac{n \sum_{i=1}^n X_1 X_2 - \sum_{i=1}^n X_1 \sum_{i=1}^n X_2}{\sqrt{\left[n \sum_{i=1}^n (X_1)^2 - \left(\sum_{i=1}^v X_1 \right)^2 \right] \left[n \sum_{i=1}^n (X_2)^2 - \left(\sum_{i=1}^v X_2 \right)^2 \right]}} \right]^2$$

$$R_j^2 = \left[\frac{(38)(35,853) - (0)(0)}{\sqrt{[(38)(32,0589) - (0)^2][(38)(37) - (0)^2]}} \right]^2$$

$$R_j^2 = (1,041)^2 = 1,08368$$

Menghitung nilai VIF :

$$VIF = \frac{1}{1 - R^2}$$
$$VIF = \frac{1}{1 - 1,08368} = -11,950$$

Menghitung nilai *tolerance* :

$$\text{Nilai } tolerance = \frac{1}{-11,950} = -0,084$$

Hasil perhitungan uji multikolineritas semua variabel terdapat pada Tabel 4.2 sebagai berikut :

Tabel 4.2 Perhitungan Uji Multikolineritas Pada Semua Variabel

Variabel	Variabel	Nilai <i>VIF</i>	Nilai <i>Tolerance</i>
x_1	x_2	16,3840	0,0610
x_1	x_3	2,8950	0,3454
x_2	x_3	2,1027	0,4756

x_4	x_6	1,0105	0,9897
x_5	x_6	1,0287	0,9721
x_1	x_7	2,1504	0,4650
x_2	x_7	2,0818	0,4804
x_3	x_7	1,5491	0,6455
x_4	x_7	1,3480	0,7419

4.4. Proseses *Hierarchical K-Means Clustering*

4.4.1. Hierarchical Clustering

1. *Single Linkage*

Data yang akan digunakan adalah data yang telah distandarisasi menggunakan *z-score* dan tidak multikolinieritas pada Lampiran J. Dibawah ini merupakan contoh dari perhitungan jarak antara objek 1 (Kab. Situbondo)

$$\begin{aligned} d_{1,2} &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_7 - y_7)^2} \\ d_{1,2} &= \sqrt{(0, 1060 - (-0, 9224))^2 + ((-0, 2023) - (-0, 5044))^2} \\ &\quad \sqrt{+(0, 3070 - (-0, 6013))^2 + \dots + ((-0, 5816) - (-1, 012))^2} \\ d_{1,2} &= 2, 4165608 \end{aligned}$$
$$\begin{aligned} d_{1,2} &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_7 - y_7)^2} \\ d_{1,2} &= \sqrt{((-0,9224) - (-0,2831))^2 + ((-0,5045) - (-0,1474))^2} \\ &\quad \sqrt{+((-0,6014) - (-0,2031))^2 + \dots + ((-1,012) - 0,2261)^2} \\ d_{1,2} &= 1,8794918 \end{aligned}$$
[illegible]

$$d_{1,2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_7 - y_7)^2}$$

$$d_{1,2} = \sqrt{((-0,2831) - 1,60404)^2 + ((-0,1474) - 0,0448)^2}$$

$$\sqrt{+((-0,2031) - 1,656)^2 + \dots + ((0,2261 - 1,4080)^2}$$

$$d_{1,2} = 4,1142566$$

Dari hasil perhitungan jarak *Euclidean* tersebut antara objek 2 (Kab. Madiun) dan objek 3 (Kab. Sumenep) didapatkan jarak sebesar 2,4165608, antara objek 2 (Kab. Madiun) dan objek 3 (Kab. Sumenep) didapatkan jarak sebesar 1,8794918 dan antara objek 3 (Kab. Sumenep) dan objek 4 (Kota Surabaya) didapatkan jarak sebesar 4,1142566. Hal itu menunjukkan bahwa jarak *Euclidean* antara Kab. Madiun dan Kab. Sumenep lebih dekat dan memiliki karakteristik yang lebih mirip dibandingkan dengan Kab. Madiun dan Kab. Situbondo. Hasil perhitungan jarak *Euclidean* antar kabupaten dan kota dapat dilihat pada Lampiran K.

Membentuk *cluster* baru dilakukan dengan cara menggabungkan jarak antara dua objek yang paling dekat atau yang mempunyai jarak terkecil. Berdasarkan Lampiran K, jarak antar kabupaten atau kota yang paling terkecil sebesar 0.39518 yaitu jarak antara *cluster* objek 2 (Kab. Madiun) dan *cluster* objek 32 (Kota Madiun), kemudian kedua objek tersebut menjadi satu *cluster*.

Kemudian dilakukan pembentukan matriks jarak baru dengan menggunakan Persamaan 2.6 berdasarkan Lampiran K. Jarak antar kabupaten atau kabupaten yang digunakan untuk pembentukan *cluster* baru adalah menggunakan objek 2 (Kab. Madiun) dan objek 32 (Kota Madiun). Perhitungan

Pada metode *Complete Linkage* proses pengelompokkannya yaitu dengan menggabungkan dua objek atau lebih yang memiliki jarak paling jauh. Untuk mengukur kedekatan atau kemiripan antar dua objek digunakan jarak *Euclidean* menggunakan Persamaan 2.5.

Berikut ini adalah perhitungan jarak *Euclidean* antara objek 1 (Kab. Situbondo) dan objek 2 (Kab. Madiun) :

Perhitungan jarak *Euclidean* antara objek 2 (Kab. Madiun) dan objek 3

Kemudian dilakukan pembentukan matriks jarak baru dengan menggunakan Persamaan 2.7 berdasarkan Lampiran K. Jarak antar kabupaten atau kabupaten yang digunakan untuk pembentukan *cluster* baru adalah menggunakan objek 2 (Kab. Madiun) dan objek 32 (Kota Madiun). Perhitungan pembentukan matriks jarak baru sebagai berikut :

$$d_{2,32}38 = \text{Max}(d_{2}38, d_{32}38) = \text{Max}(d_{1,224}, d_{0,907}) = 1,224$$

Perhitungan dilanjutkan hingga matriks jarak dari semua objek yang telah digabungkan dan mendapatkan matriks baru. Jumlah *cluster* dan anggota dengan metode *complete linkage* pada bulan Desember 2020 pada Lampiran M. Proses ini dapat digambarkan dalam bentuk dendogram. Berikut adalah dendogram pengelompokkan kabupaten atau kota di provinsi Jawa Timur menggunakan metode *complete linkage* jika ingin membentuk 2 *cluster* pada Gambar 4.4 :

Berikut ini adalah perhitungan jarak *Euclidean* antara objek 1 (Kab.

ngan diatas menggunakan pengelompokkan dengan
dian dilanjut dengan metode *complete linkage* dan
4.4 merupakan hasil pusat *cluster* pada percobaan
linkage :

Tabel 4.4 Pusat *cluster* (*centroid*) pada 2 *cluster single linkage*

	Konfirmasi	Meninggal	Bergejala	Tp.Gejala	Perjalanan	Kontak	Tp.Riwayat
C1	0,0184	0,0025	0,0034	0,0330	-0,1452	0,0251	0,0332
C2	-0,6818	-0,0925	-0,1274	-1,223	5,372	-0,9305	-1,229

Kab. Pamekasan	1,565	6,148	1
Kab. Bojonegoro	1,618	6,174	1
Kota Mojokerto	1,828	5,898	1
Kab. Lamongan	1,696	6,659	1
Kab. Trenggalek	1,078	5,709	1
Kab. Pasuruan	1,757	6,162	1
Kota Blitar	1,459	5,911	1
Kab. Pacitan	1,231	5,856	1
Kab. Banyuwangi	5,585	8,585	1
Kab. Kediri	3,614	8,071	1
Kota Kediri	5,920	0,000	2
Kab. Ponorogo	1,329	6,257	1
Kab. Blitar	1,459	6,334	1
Kab. Sampang	2,185	5,487	1
Kota Malang	4,642	8,485	1
Kab. Jember	7,043	10,684	1
Kab. Sidoarjo	1,669	6,402	1
Kab. Tulungagung	1,539	5,009	1
Kota Batu	1,915	5,917	1
Kab. Jombang	2,041	6,609	1
Kab. Ngawi	1,584	5,383	1
Kab. Nganjuk	1,326	5,509	1
Kota Probolinggo	1,589	5,431	1
Kab. Tuban	2,706	7,228	1
Kab. Mojokerto	1,734	5,960	1

$$\begin{aligned}
C1_{Konfirmasi} &= \frac{0,106 + (-0,922) + (-0,283) + \dots + 0,021 + (-0,707)}{37} = 0,0184 \\
C1_{Meninggal} &= \frac{(-0,202) + (-0,504) + (-0,147) + \dots + 0,504}{37} = 0,0025 \\
C1_{Bergejala} &= \frac{0,307 + (-0,601) + (-0,203) + \dots + (-0,795)}{37} = 0,0034 \\
C1_{TanpaGejala} &= \frac{(-0,236) + (-1,018) + (-0,285) + \dots + (-0,285)}{37} = 0,0330 \\
C1_{Perjalanan} &= \frac{(-0,055) + 0,297 + (-0,447) + \dots + (-0,447)}{37} = -0,1452 \\
C1_{Kontak} &= \frac{1,261 + (-0,454) + (-0,914) + \dots + 0,108}{37} = 0,0251 \\
C1_{TanpaRiwayat} &= \frac{(-0,581) + (-1,012) + 0,226 + \dots + (-0,933)}{37} = 0,03322
\end{aligned}$$

$$\begin{aligned} C2_{Konfirmasi} &= \frac{-0,681}{1} = -0,681 \\ C2_{Meninggal} &= \frac{-0,092}{1} = -0,092 \\ C2_{Bergejala} &= \frac{-0,127}{1} = -0,127 \\ C2_{TanpaGejala} &= \frac{-1,223}{1} = -1,223 \\ C2_{Perjalanan} &= \frac{(5,372}{1} = 5,372 \\ C2_{Kontak} &= \frac{-0,930}{1} = -0,930 \\ C2_{TanpaRiwayat} &= \frac{-1,229}{1} = -1,229 \end{aligned}$$

Hasil perhitungan pusat *cluster* (*centroid*) dari data yang masuk *cluster* yang sama akan dipergunakan untuk membentuk *centroid* baru. Hasil pembentukan *centroid* baru ada pada Tabel 4.6

Tabel 4.6 Pusat *cluster* (*centroid*) pada 2 *cluster single linkage*

	Konfirmasi	Meninggal	Bergejala	Tp.Gejala	Perjalanan	Kontak	Tp.Riwayat
C1	0,0184	0,0025	0,0034	0,0330	-0,1452	0,0251	0,0332
C2	-0,6818	-0,0925	-0,1274	-1,223	5,372	-0,9305	-1,229

Dari Tabel 4.6 tidak terdapat data yang berubah pusat *cluster* maka iterasi berhenti. Dari hasil *clustering* yang diperoleh berhenti pada iterasi ke-1 dengan hasil *centroid* pada Tabel 4.6. Perhitungan diatas menggunakan pengelompokkan dengan metode *single Linkage*, perhitungan kemudian dilanjut dengan metode *complete linkage* dan *average linkage*. Jumlah *cluster* dan anggota menggunakan metode *HK-Means single linkage* ada pada Lampiran O. Jumlah *cluster* dan anggota menggunakan metode *HK-Means complete linkage* ada pada Lampiran P. Dan jumlah *cluster* dan anggota menggunakan metode *HK-Means single linkage* ada

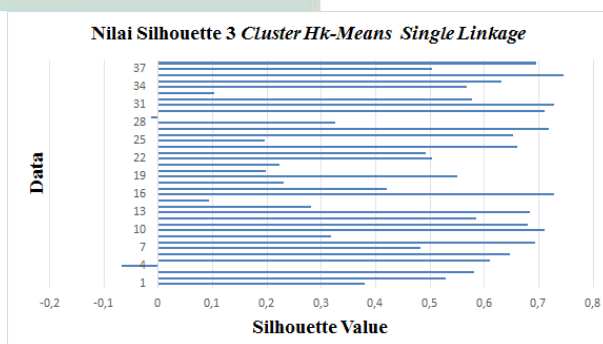
4.5. Pemilihan Metode Terbaik Dengan *Sillhoutte Coefficient*

Tabel 4.7 Nilai *Sillhoutte Coefficient* masing-masing percobaan *cluster*

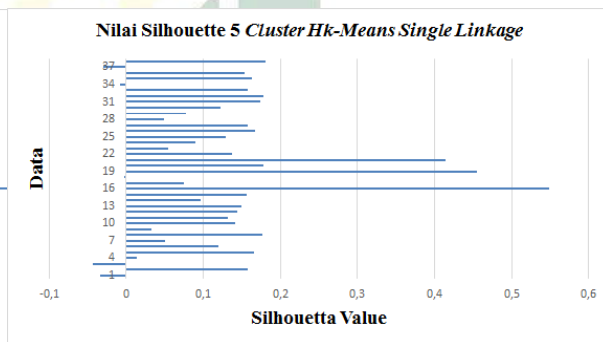
Metode	Percobaan <i>Cluster</i>	<i>Sillhoutte Coefficient</i>
<i>Single Linkage</i>	<i>2 Cluster</i>	0,5533
	<i>3 Cluster</i>	0,4831
	<i>4 Cluster</i>	0,1670
	<i>5 Cluster</i>	0,1338
<i>Complete Linkage</i>	<i>2 Cluster</i>	0,5628
	<i>3 Cluster</i>	0,4983
	<i>4 Cluster</i>	0,1647
	<i>5 Cluster</i>	0,3289
<i>Average Linkage</i>	<i>2 Cluster</i>	0,5533
	<i>3 Cluster</i>	0,4983
	<i>4 Cluster</i>	0,1647
	<i>5 Cluster</i>	0,1368

[illegible]

Nilai *silhouette* dari masing-masing data tiap percobaan *cluster* akan disajikan dalam bentuk gambar grafik pada Gambar 4.6 - 4.17.

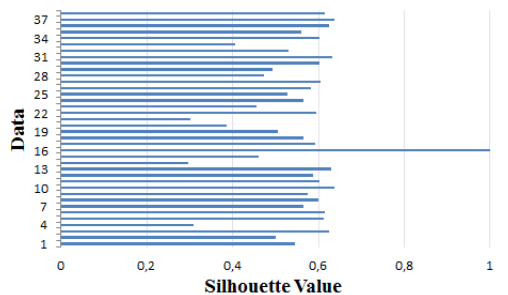


Gambar 4.7 HK-Means Single Linkage 3 cluster

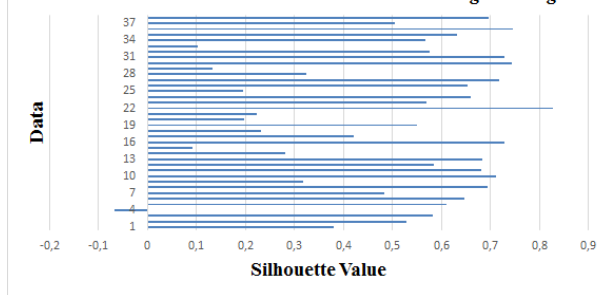


Gambar 4.9 *HK-Means Single Linkage 5 cluster*

Gambar 4.6-4.9 menunjukkan keakuratan jumlah setiap *cluster*, sehingga diketahui manakah *cluster* terbaik untuk struktur data yang kuat. Untuk menentukan *cluster* terbaik menggunakan *silhouette* dilihat grafik pada Gambar 4.6-4.9. Jika pada grafik terdapat hasil *cluster* yang berada disebelah kiri 0 artinya nilai *silhouette* kurang dari 0 atau minus. Hal tersebut menandakan data *overlapping* atau tidak

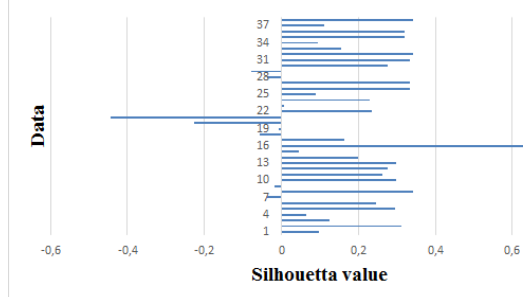
Nilai Silhouette 2 Cluster Hk-Means Single Linkage

Gambar 4.14 *HK-Means Average Linkage 2 cluster*

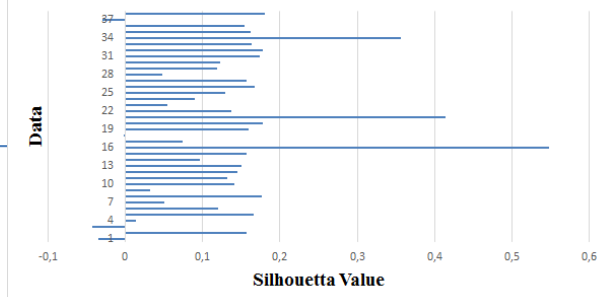
Nilai Silhouette 3 Cluster Hk-Means Average Linkage

Gambar 4.15 *HK-Means Average Linkage 3 cluster*

Nilai Silhouette 4 Cluster Hk-Means Average Linkage



Gambar 4.16 *HK-Means Average Linkage 4 cluster*

Nilai Silhouette 5 Cluster Hk-Means Average Linkage

Gambar 4.17 HK-Means Average Linkage 5 cluster

Dilihat pada Grafik yang ada pada Gambar 4.14-4.17 terdapat data yang memiliki nilai *silhouette* negatif atau *overlapping*, kecuali Gambar 4.14 tidak terdapat data yang memiliki nilai *silhouette* negatif atau *overlapping*. Hal tersebut membuat hasil keakuratannya tidak bagus.

Berdasarkan nilai *silhouette coefficient* menunjukkan percobaan yang memiliki tingkat kekauratan lebih baik ada di percobaan 2 *cluster* dengan metode *complete linkage* dengan nilai 0,5628 dan tidak terdapat data yang memiliki nilai *silhouette* negatif atau *overlapping*.

4.6. Interpretasi *Cluster* Optimal

Pada analisis *clustering* menggunakan *HK-Means complete linkage* didapat pusat *cluster/centroid* pada data yang telah dinormalisasi yaitu data yang telah

Pada Tabel 4.8 dapat dianalisis bahwa setiap *centroid* memiliki pola tertentu. Hal ini digambarkan berdasarkan pada pola nilai *centroid* seperti pada Tabel 4.9.

Tabel 4.9 Centroid Akhir pada 2 cluster Complete linkage

[illegible]

Pada Tabel 4.9 dapat disimpulkan bahwa untuk C1 dikategorikan sebagai jenis *cluster* aman (zona hijau) karena memiliki pola *centroid* yang nilainya rendah. Dan C2 dikategorikan sebagai jenis *cluster* rawan (zona merah) karena memiliki pola *centroid* yang nilainya tinggi. Berikut adalah kelompok Kabupaten atau kota berdasarkan hasil analisis *cluster* menggunakan *HK-Means(Complete Linkage)* :

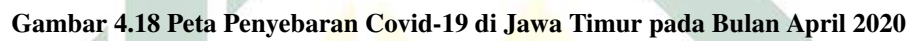
Pada Tabel diatas terdapat 6 kabupaten atau kota yaitu Kota Surabaya, Kab. Jember , Kab. Kediri, Kab. Tuban, Kab. Banyuwangi, Kota Malang berada pada kategori daerah rawan Covid-19 (zona merah). Sedangkan 32 kabupaten atau kota lainnya pada kategori aman (zona hijau).

Setelah melakukan *HK-Means clustering* dan validasi *cluster*, selanjutnya yaitu pemetaan daerah kabupaten atau kota di Jawa Timur berdasarkan analisis *cluster* yaitu pada Tabel 4.12 : H = Hijau , M = Merah

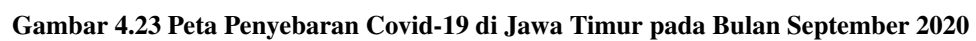
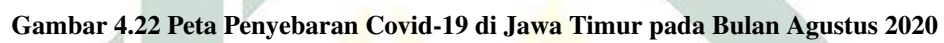
Tabel 4.12 Pemetaan Kabupaten/Kota Berdasarkan Hasil Analisis *Cluster*

Kab/Kota	Apr	Mei	Juni	Juli	Agt	Sep	Okt	Nop	Des
Kab. Situbondo	H	H	H	H	H	H	H	H	H
Kab. Madiun	H	H	H	H	H	H	H	H	H
Kab. Sumenep	H	H	H	H	H	H	H	H	H
Kota Surabaya	M	M	M	M	M	M	M	M	M
Kab. Gresik	H	H	H	H	H	H	H	H	H
Kab. Pamekasan	H	H	H	H	H	H	H	H	H
Kab. Bojonegoro	H	H	H	H	H	H	H	H	H
Kota Mojokerto	H	H	H	H	H	H	H	H	H
Kab. Lamongan	H	H	H	H	H	H	H	H	H
Kab. Trenggalek	H	H	H	H	H	H	H	H	H
Kab. Pasuruan	H	H	H	H	H	H	H	H	H
Kota Blitar	H	H	H	H	H	H	H	H	H
Kab. Pacitan	H	H	H	H	H	H	H	H	H
Kab. Banyuwangi	H	H	H	H	H	H	H	M	M

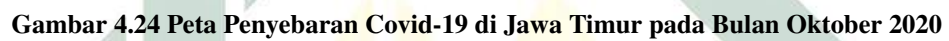
Kab/Kota	Apr	Mei	Juni	Juli	Agt	Sep	Okt	Nop	Des
Kab. Kediri	H	H	H	H	H	H	H	H	M
Kota Kediri	H	H	H	H	H	H	H	H	H
Kab. Ponorogo	H	H	H	H	H	H	H	H	H
Kab. Blitar	H	H	H	H	H	H	H	H	H
Kab. Sampang	H	H	H	H	H	H	H	H	H
Kota Malang	H	H	H	H	H	H	H	H	M
Kab. Jember	H	H	H	H	H	H	H	M	M
Kab. Sidoarjo	H	H	H	H	H	M	M	H	H
Kab. Tulungagung	H	H	H	H	H	H	H	H	H
Kota Batu	H	H	H	H	H	H	H	H	H
Kab. Jombang	H	H	H	H	H	H	H	M	H
Kab. Ngawi	H	H	H	H	H	H	H	H	H
Kab. Nganjuk	H	H	H	H	H	H	H	H	H
Kota Probolinggo	H	H	H	H	H	H	H	H	H
Kab. Tuban	H	H	H	H	H	H	H	H	M
Kab. Mojokerto	H	H	H	H	H	H	H	H	H
Kab. Magetan	H	H	H	H	H	H	H	H	H
Kota Madiun	H	H	H	H	H	H	H	H	H
Kab. Lumajang	H	H	H	H	H	H	H	M	H
Kab. Bondowoso	H	H	H	H	H	H	H	H	H
Kota Pasuruan	H	H	H	H	H	H	H	H	H
Kab. Malang	H	H	H	H	H	H	H	H	H
Kab. Probolinggo	H	H	H	H	H	H	H	H	H
Kab. Bangkalan	H	H	H	H	H	H	H	H	H



Gambar 4.19 Peta Penyebaran Covid-19 di Jawa Timur pada Bulan Mei 2020



Tetapi pada bulan September 2020, dapat dilihat pada Gambar 4.23 bahwa daerah yang rawan Covid-19 bertambah yaitu menjadi Kota Surabaya dan, Kab. Sidoarjo.



Gambar 4.25 Peta Penyebaran Covid-19 di Jawa Timur pada Bulan Nopember 2020

Kota Surabaya menjadi satu-satunya yang berada pada daerah yang rawan (Zona merah) Covid-19 dari bulan April-Desember 2020. Hal tersebut dipengaruhi oleh beberapa faktor yaitu, karena kota Surabaya merupakan ibukota provinsi Jawa Timur sehingga penduduknya banyak dan aktivitas masyarakatnya juga tinggi. Beberapa faktor lainnya yaitu karena banyak masyarakatnya yang melakukan aktivitas tanpa mematuhi protokol kesehatan. Dan faktor penyebab lonjakan kasus positif lainnya di Kota Surabaya adalah karena budaya dan karakteristik masyarakatnya, masyarakat Surabaya terkenal dengan sifat yang keras dan berani. Dan pada masyarakat Surabaya terdapat budaya "cangkruan". Cangkruan atau cangkruk adalah kegiatan yang digunakan sebagai tempat pusat komunikasi, sosialisasi, informasi dan hiburan. Budaya cangkruk ini telah menjadi gaya hidup banyak masyarakat Surabaya. Dengan budaya sehingga memunculkan banyak kerumunan yang berdampak pada bertambahnya kasus Covid-19 di Kota Surabaya. (Apriliyanti et.al, 2021)

Ditengah pandemi wabah Covid-19 saat ini, untuk menjaga kesehatan dan keselamatan diri dan keluarga kita sangat penting. Dengan mematuhi protokol ke-

Maka dalam situasi ini masyarakat yang berada dalam *cluster* merah atau daerah rawan dilarang untuk keluar atau berpergian terutama berpergian ke daerah dalam *cluster* hijau atau daerah aman. Karena dapat menyebarkan atau membawa virus Covid-19 agar virus Covid-19 yang ada di daerahnya tidak meluas ke daerah lain. Begitupun sebaliknya masyarakat yang berada dalam *cluster* hijau atau daerah aman juga dilarang untuk berpergian atau memasuki daerah *cluster* merah atau

daerah rawan untuk menghindari tertular virus Covid-19.

Sudah menjadi kodrat manusia bahwa dalam menjalani kehidupan di bumi ini tidak terlepas dari cobaan, musibah salah satunya contohnya seperti adanya wabah Covid-19 ini. Allah SWT berfirman pada surat Al-Baqarah ayat 155 :

وَلَنَبْلُوَنَّكُمْ بِشَيْءٍ مِّنَ الْخَوْفِ وَالْجُوعِ وَنَقْصٍ مِّنَ الْأَمْوَالِ وَالْأَنْفُسِ وَالثَّمَرَاتِ ۗ وَبَشِّرِ الصَّابِرِينَ

Artinya : "Dan sungguh akan kami berikan cobaan kepadamu, dengan sedikit ketakutan, kelaparan, kekurangan harta, jiwa dan buah-buahan. Dan berikanlah berita gembira kepada orang-orang yang sabar". Maka bersabar merupakan ikhtiar dan penolong bagi manusia untuk lulus dari musibah Covid-19, seperti dijelaskan pada surat Al-Baqarah ayat 45, yang mempunyai arti : "Jadikanlah sabar dan salat sebagai penolongmu. Dan sesungguhnya yang demikian itu sungguh berat, kecuali bagi orang-orang yang khusyu".

Oleh karena itu bersabar merupakan ikhtiar bagi kita untuk lulus dari musibah pandemi Covid-19. Jadi, sebagai manusia selain kewajiban menjaga kesehatan, tetapi mereka juga harus bersabar dalam menghadapi musibah pandemi ini.

PENUTUP

5.1. Simpulan

1. Metode *HK-Means* yang menghasilkan hasil *cluster* dengan struktur terbaik dilihat dari nilai *Silhouette Coefficient* pada data Covid-19 di Jawa Timur adalah metode *HK-Means Complete Linkage* dengan nilai *Silhouette* 0,5628. Hasil *cluster* ini termasuk dalam struktur *cluster* standar karena nilai *silhouette* nya lebih dari 0,5 dan kurang dari sama dengan 0,7. Dan tidak terdapat data yang bernilai negatif atau *overlapping*. Dengan nilai pusat *cluster* pada C1 konfirmasi -0,259, C1 meninggal -0,357, C1 bergejala -0,290, C1 tanpa gejala -0,313, C1 perjalanan -0,021, C1 kontak -0,300, dan C1 tanpa riwayat -0,300. Sedangkan nilai pusat *cluster* pada C2 konfirmasi 1,916, C2 meninggal 1,848, C2 bergejala 1,549, C2 tanpa gejala 1,670, C2 perjalanan 0,114, C2 kontak 1,602, dan C2 tanpa riwayat 1,601.
2. Hasil *cluster* yang terbentuk menggunakan metode terbaik dari *HK-Means complete linkage* pada data Covid-19 di Jawa Timur pada bulan Desember 2020 menghasilkan 32 kabupaten atau kota berada pada *cluster* 1 (C1) atau

Untuk menghasilkan hasil yang lebih bervariasi dan juga sebagai pembandingan dalam proses *clustering*, penulis ingin menyampaikan beberapa saran.

- [illegible]

M Tirta Raharja menggunakan Algoritma K-Mean

ulshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomay

(2014). A survey of clustering algorithms for big

al analysis,” IEEE Trans. Emerg. Top. Comput., v

tos, B., Barakbah, A. R. (2012). Analisa Perba

Clustering, K-means dan Gabungan Keduanya dal

(1), 521–525. <https://doi.org/10.12962/j23373539.v>

h, T. Y., Saybani, M. R., dan Yazdi, S. R. A. S.

rid based clustering algorithms on data streams. Da

n fuzzy systems and knowledge discovery (fskd) (h

- 79

<http://jatimprov.go.id>. (2015). Profil Provinsi Jawa Timur. Diakses 17 September 2020, dari <http://jatimprov.go.id/read/profil>.

<https://covid19.go.id>. (2020). Tanya Jawab tentang COVID-19. Diakses 17 September 2020, dari <https://covid19.go.id/tanya-jawab>.

http://infocovid19.jatimprov.go.id. (2020). Peta Sebaran COVID-19 Jatim. Diakses 17 September 2020, dari <http://infocovid19.jatimprov.go.id/>.

Laeli, S.(2014). Analisis Cluster dengan Average Linkage Method dan Wards Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link. Yogyakarta : Universitas Negeri Yogyakarta.

Latan, H. (2014). Aplikasi analisis data statistik untuk ilmu sosial sains dengan STATA. Bandung: Alfabeta.

Mahmudah, A.(2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. Jurnal Matematika Statistika Dan Komputasi, 16(2), 174–186. <https://doi.org/10.20956/jmsk.v>

Ningsih, S., Wahyuningsih, S., Nasution, N.Y. (2016). Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Hasil Analisis Cluster (Studi Kasus Produksi Palawija Provinsi Kalimantan Timur 2014=2015). Prosiding Seminar Sains dan Teknologi FMIPA Unmul, 01, 01.

- ...